Meta-analysis of the heritability of human traits based on fifty years of twin studies

Tinca J C Polderman[1,10], Beben Benyamin[2,10], Christiaan A de Leeuw[1,3], Patrick F Sullivan[4–6], Arjen van Bochoven[7], Peter M Visscher[2,8,11] & Danielle Posthuma[1,9,11]

Despite a century of research on complex traits in humans, the relative importance and specific nature of the influences of genes and environment on human traits remain controversial. We report a meta-analysis of twin correlations and reported variance components for 17,804 traits from 2,748 publications including 14,558,903 partly dependent twin pairs, virtually all published twin studies of complex traits. Estimates of heritability cluster strongly within functional domains, and across all traits the reported heritability is 49%. For a majority (69%) of traits, the observed twin correlations are consistent with a simple and parsimonious model where twin resemblance is solely due to additive genetic variation. The data are inconsistent with substantial influences from shared environment or non-additive genetic variation. This study provides the most comprehensive analysis of the causes of individual differences in human traits thus far and will guide future gene-mapping efforts. All the results can be visualized using the MaTCH webtool.

Insight into the nature of observed variation in human traits is important in medicine, psychology, social sciences and evolutionary biology. It has gained new relevance with both the ability to map genes for human traits and the availability of large, collaborative data sets to do so on an extensive and comprehensive scale. Individual differences in human traits have been studied for more than a century, yet the causes of variation in human traits remain uncertain and controversial.

Specifically, the partitioning of observed variability into underlying genetic and environmental sources and the relative importance of additive and non-additive genetic variation are continually debated[1–5]. Recent results from large-scale genome-wide association studies (GWAS) show that many genetic variants contribute to the variation in complex traits and that effect sizes are typically small[6,7]. However, the sum of the variance explained by the detected variants is much smaller than the reported heritability of the trait[4,6–10]. This 'missing heritability' has led some investigators to conclude that non-additive variation must be important[4,11]. Although the presence of gene-gene interaction has been demonstrated empirically[5,12–17], little is known about its relative contribution to observed variation[18].

In this study, our aim is twofold. First, we analyze empirical estimates of the relative contributions of genes and environment for virtually all human traits investigated in the past 50 years. Second, we assess empirical evidence for the presence and relative importance of non-additive genetic influences on all human traits studied. We rely on classical twin studies, as the twin design has been used widely to disentangle the relative contributions of genes and environment, across a variety of human traits. The classical twin design is based on contrasting the trait resemblance of monozygotic and dizygotic twin pairs. Monozygotic twins are genetically identical, and dizygotic twins are genetically full siblings. We show that, for a majority of traits (69%), the observed statistics are consistent with a simple and parsimonious model where the observed variation is solely due to additive genetic variation. The data are inconsistent with a substantial influence from shared environment or non-additive genetic variation. We also show that estimates of heritability cluster strongly within functional domains, and across all traits the reported heritability is 49%. Our results are based on a meta-analysis of twin correlations and reported variance components for 17,804 traits from 2,748 publications including 14,558,903 partly dependent twin pairs, virtually all twin studies of complex traits published between 1958 and 2012. This study provides the most comprehensive analysis of the causes of individual differences in human traits thus far and will guide future gene-mapping efforts. All results can be visualized with the accompanying MaTCH webtool.

RESULTS

The distribution of studied traits is nonrandom

We systematically retrieved published classical twin studies in which observed variation in human traits was partitioned into genetic and environmental influences. For each study, we collected reported

[1]Department of Complex Trait Genetics, VU University, Center for Neurogenomics and Cognitive Research, Amsterdam, the Netherlands. [2]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [3]Institute for Computing and Information Sciences, Radboud University Nijmegen, Nijmegen, the Netherlands. [4]Center for Psychiatric Genomics, Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA. [5]Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina, USA. [6]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [7]Faculty of Sciences, VU University, Amsterdam, the Netherlands. [8]University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia. [9]Department of Clinical Genetics, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, the Netherlands. [10]These authors contributed equally to this work. [11]These authors jointly supervised this work. Correspondence should be addressed to D.P. (d.posthuma@vu.nl).

twin correlations for continuous traits and contingency tables for dichotomous traits, estimates from genetic model-fitting and study characteristics (sample size, population, age cohort and ascertainment scheme) (**Supplementary Table 1** and **Supplementary Note**). We manually classified the investigated traits using the chapter and subchapter levels of the International Classification of Functioning, Disability and Health (ICF) or the International Statistical Classification of Diseases and Related Health Problems (ICD-10) (Online Methods). ICD-10 and ICF subchapter levels refer to actual diseases (for example, atopic dermatitis) and traits (for example, temperament and personality functions), respectively. We identified 2,748 relevant twin studies, published between 1958 and 2012. Half of these were published after 2004, with sample sizes per study in 2012 of around 1,000 twin pairs (**Supplementary Table 2**). Each study could report on multiple traits measured in one or several samples. These 2,748 studies reported on 17,804 traits. Twin subjects came from 39 different countries, with a large proportion of studies (34%) based on US twin samples. The continents of South America (0.5%), Africa (0.2%) and Asia (5%) were heavily underrepresented (**Fig. 1a,b** and **Supplementary Table 3**). The total number of studied twins was 14,558,903 partly dependent pairs, or 2,247,128 when correcting for reporting on multiple traits per study. The majority

of studies (59%) were based on the adult population (aged 18–64 years), although the sample sizes available for studies of the elderly population (aged 65 years or older) were the largest (**Supplementary Table 4**). Authorship network analyses showed that 61 communities of authors wrote the 2,748 published studies. The 11 largest authorship communities contained >65 authors and could be mapped back to the main international twin registries, such as the Vietnam Era Twin Registry, the Finnish Twin Cohort and the Swedish Twin Registry (**Supplementary Fig. 1**).

The investigated traits fell into 28 general trait domains. The distribution of the traits evaluated in twin studies was highly skewed, with 51% of studies focusing on traits classified under the psychiatric, metabolic and cognitive domains, whereas traits classified under the developmental, connective tissue and infection domains together accounted for less than 1% of all investigated traits (**Fig. 1c** and **Supplementary Tables 5–7**). The ten most investigated traits were temperament and personality functions, weight maintenance functions, general metabolic functions, depressive episode, higher-level cognitive functions, conduct disorders, mental and behavioral disorders due to use of alcohol, anxiety disorders, height and mental and behavioral disorders due to use of tobacco. Collectively, these traits accounted for 59% of all investigated traits.
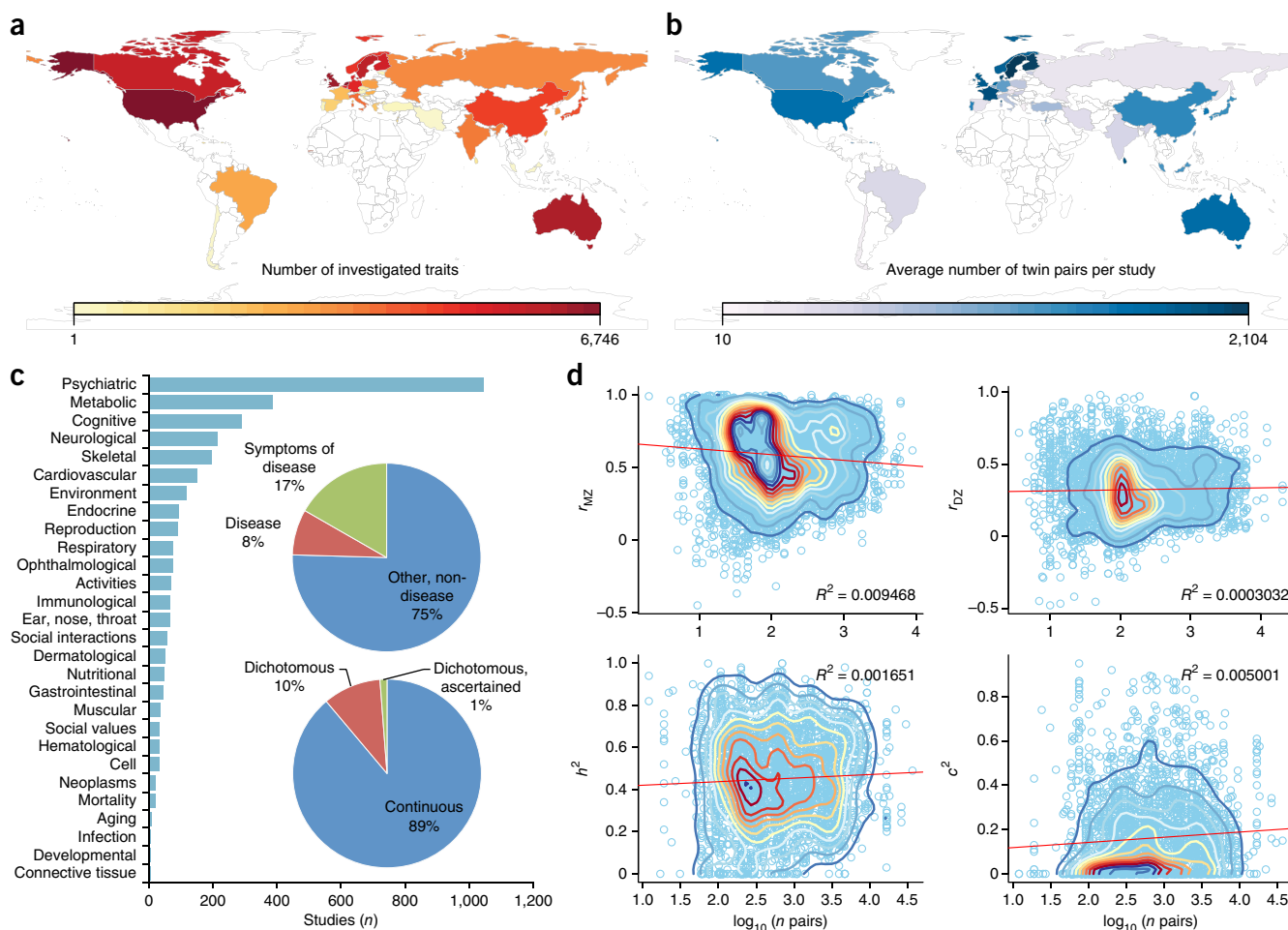

Figure 1 Distribution of the investigated traits in virtually all twin studies published between 1958 and 2012. (**a**) The number of investigated traits in classical twin studies across all countries. (**b**) The average number of twin pairs included per study across countries. (**c**) The number of investigated traits according to functional trait domain and trait characteristic (inset). (**d**) Monozygotic and dizygotic twin correlations and reported estimates of h^2 and c^2 as a function of sample size. Contour lines indicate the density of the data in that region. The lines are colored by 'heat' from blue to red, indicating increasing data density.
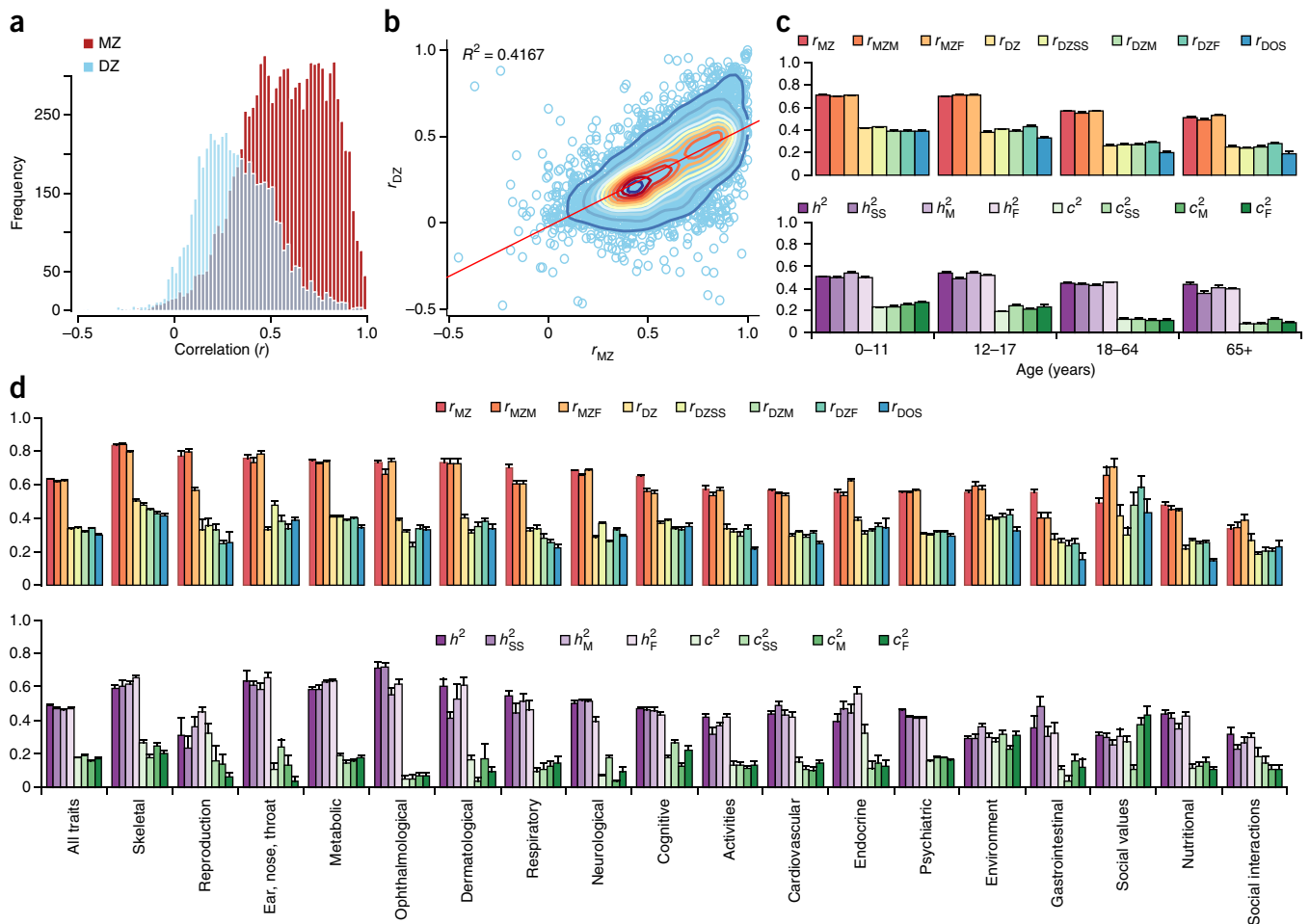
Figure 2 Twin correlations and heritabilities for all human traits studied. (**a**) Distribution of r_{MZ} and r_{DZ} estimates across the traits investigated in 2,748 twin studies published between 1958 and 2012. r_{MZ} estimates are based on 9,568 traits and 2,563,628 partly dependent twin pairs; r_{DZ} estimates are based on 5,220 traits and 2,606,252 partly dependent twin pairs (**Table 1**). (**b**) Relationship between r_{MZ} and r_{DZ}, using all 5,185 traits for which both were reported. (**c**) Random-effects meta-analytic estimates of twin correlations (top) and reported variance components (bottom) across all traits separately for four age cohorts. Error bars, standard errors. (**d**) Random-effects meta-analytic estimates of twin correlations (top) and reported variance components (bottom) across all traits, and within functional domains for which data on all correlations and variance components were available. Error bars, standard errors.

Equal contribution of genes and environment

We did not find evidence of systematic publication bias as a function of sample size (for example, where studies based on relatively small samples were only published when larger effects were reported) (**Fig. 1d**, **Supplementary Figs. 2–6** and **Supplementary Tables 8–11**). We calculated the weighted averages of correlations for monozygotic (r_{MZ}) and dizygotic (r_{DZ}) twins and of the reported estimates of the relative contributions of genetic and environmental influences to the investigated traits using a random-effects meta-analytic model to allow for heterogeneity across different studies (**Supplementary Tables 12–15**). The meta-analyses of all traits yielded an average r_{MZ} of 0.636 (s.e.m. = 0.002) and an average r_{DZ} of 0.339 (s.e.m. = 0.003). The reported heritability (h^2) across all traits was 0.488 (s.e.m. = 0.004), and the reported estimate of shared environmental effects (c^2) was 0.174 (s.e.m. = 0.004) (**Fig. 2a**,**b**, **Table 1** and **Supplementary Fig. 7**).

Estimates of h^2 and c^2 cluster in functional domains

We found that heritability estimates clustered in functional domains, with the largest heritability estimates for traits classified under the ophthalmological domain ($h^2 = 0.712$, s.e.m. = 0.041), followed by the ear, nose and throat ($h^2 = 0.637$, s.e.m. = 0.064), dermatological ($h^2 = 0.604$, s.e.m. = 0.043) and skeletal ($h^2 = 0.591$, s.e.m. = 0.018) domains. The lowest heritability estimates were for traits in the environment, reproduction and social values domains (**Fig. 2d** and **Supplementary Table 16**). All weighted averages of h^2 across >500 distinct traits had a mean greater than zero (**Supplementary Tables 17–24**). The lowest reported heritability for a specific trait was for gene expression, with an estimated $h^2 = 0.055$ (s.e.m. = 0.026) and an estimated c^2 of 0.736 (s.e.m. = 0.033) (but note that these trait averages are based on reported estimates of variance components derived from only 20 data points reporting on the expression levels of 20 genes; **Supplementary Table 21**). We found the largest influence of c^2 for traits in the cell domain ($c^2 = 0.674$, s.e.m. = 0.048), followed by traits in the infection ($c^2 = 0.351$, s.e.m. = 0.153), hematological ($c^2 = 0.324$, s.e.m. = 0.090), endocrine ($c^2 = 0.322$, s.e.m. = 0.050), reproduction ($c^2 = 0.320$, s.e.m. = 0.061), social values ($c^2 = 0.271$, s.e.m. = 0.032), environment ($c^2 = 0.269$, s.e.m. = 0.020) and skeletal ($c^2 = 0.265$, s.e.m. = 0.019) domains (**Fig. 2d** and **Supplementary Table 16**).

Table 1 Weighted means of twin correlations and variance components across all human traits investigated in a classical twin study and published between 1958 and 2012

Statistic	Estimate (s.e.m.)	n traits	n pairs
r_{MZ}	0.636 (0.002)	9,568	2,563,627
r_{MZM}	0.617 (0.004)	4,518	1,070,962
r_{MZF}	0.626 (0.004)	4,360	1,171,841
r_{DZ}	0.339 (0.003)	5,220	2,606,252
r_{DZSS}	0.345 (0.003)	6,108	1,752,952
r_{DZM}	0.321 (0.003)	4,412	1,039,238
r_{DZF}	0.342 (0.004)	4,255	1,068,562
r_{DOS}	0.302 (0.005)	2,342	898,610
h^2	0.488 (0.004)	2,929	4,341,721
h^2 (same sex)	0.471 (0.005)	1,795	1,187,837
h^2 (male)	0.465 (0.005)	2,095	1,732,622
h^2 (female)	0.472 (0.005)	1,957	1,539,582
c^2	0.174 (0.004)	2,771	4,272,318
c^2 (same sex)	0.189 (0.005)	1,769	1,185,116
c^2 (male)	0.157 (0.004)	1,988	1,519,148
c^2 (female)	0.169 (0.005)	1,925	1,516,192
$2(r_{MZ} - r_{DZ})$	0.593 (0.008)	9,568	5,169,879
$2(r_{MZ} - r_{DZ})$ (same sex)	0.581 (0.008)	9,568	4,316,578
$2(r_{MZ} - r_{DZ})$ (male)	0.593 (0.010)	4,518	2,110,200
$2(r_{MZ} - r_{DZ})$ (female)	0.569 (0.010)	4,360	2,240,403
$2r_{DZ} - r_{MZ}$	0.042 (0.007)	9,568	5,169,879
$2r_{DZ} - r_{MZ}$ (same sex)	0.055 (0.006)	9,568	4,316,578
$2r_{DZ} - r_{MZ}$ (male)	0.025 (0.008)	4,518	2,110,200
$2r_{DZ} - r_{MZ}$ (female)	0.057 (0.008)	4,360	2,240,403

r, correlation; MZ, monozygotic twins; DZ, dizygotic twins; MZM, monozygotic twins, male; MZF, monzygotic twins, female; DZSS, dizygotic twins, same sex; DZM, dizygotic twins, both male; DZF, dizygotic twins, both female; DOS, dizygotic twins, opposite sex; h^2, heritability; c^2, proportion of variance due to shared environmental variation; estimate, estimate based on random-effects meta-analysis; n traits, number of investigated traits; n pairs, number of dependent twin pairs. The pairs are not independent, as the same or an overlapping sample of twins may have been used for multiple traits and across multiple studies.

Heterogeneity of twin correlations across sex and age

Across all traits, the weighted averages of twin correlations and reported h^2 and c^2 values did not show evidence of heterogeneity by sex, although there was some evidence for lower correlation in opposite-sex twin pairs in comparison to same-sex dizygotic twin pairs (**Table 1** and **Supplementary Note**). The data showed a decrease in monozygotic and dizygotic twin resemblance after adolescence and an accompanying decrease in the estimates of both h^2 and c^2 (**Fig. 2c** and **Supplementary Table 15**).

In the top 20 most investigated traits for twin correlations, the weighted estimates did not show consistent evidence for heterogeneity by sex, with r_{MZM} (monozygotic male twin correlation) and r_{MZF} (monozygotic female twin correlation) as well as r_{DZM} (dizygotic male twin correlation) and r_{DZF} (dizygotic female twin correlation) remarkably similar across the majority of the top 20 specific traits investigated (**Fig. 3**), although for several traits the correlations for opposite-sex twin pairs were lower than the estimates for same-sex twin pairs, mostly after age 11 years (for example, for weight maintenance functions, functions of brain, mental and behavioral disorders due to the use of alcohol, and mental and behavioral disorders due to the use of tobacco). Heterogeneity of weighted twin correlations by age was more prominent than heterogeneity by sex (**Fig. 3**). For example, when considering r_{MZ}, we note that for most of the top 20 investigated traits the estimate tended to decrease with age, especially after adolescence, a trend that was generally mirrored

in the r_{DZ} estimates (**Fig. 3**). Meta-analysis results across all traits across different countries are provided in **Supplementary Table 25**.

Model-fitting and selection leads to underestimation of h^2

Falconer's equations can be used to calculate $\hat{h}^2$ and $\hat{c}^2$ on the basis of twin correlations[18]. In these equations, $\hat{h}^2 = 2 \times (r_{MZ} - r_{DZ})$ and $\hat{c}^2 = 2 \times r_{DZ} - r_{MZ}$. When these are applied using the weighted averages of r_{MZ} and r_{DZ}, we find an $\hat{h}^2$ estimate of $2 \times (0.636 - 0.339) = 0.593$ and a $\hat{c}^2$ estimate of $2 \times 0.339 - 0.636 = 0.042$ (**Table 1** and **Supplementary Fig. 8**). We note that the $\hat{h}^2$ estimate based on twin correlations is larger than the weighted average of reported h^2 values (**Supplementary Figs. 9** and **10**). As a consequence, the $\hat{c}^2$ estimate based on twin correlations is lower than the weighted average of the reported c^2 component. To test whether this discrepancy was due to a bias in studies reporting only twin correlations or only variance components, we conducted the meta-analysis only on studies reporting both. This analysis yielded similar estimates with a similar discrepancy (**Supplementary Table 13**), ruling out the explanation that twin correlations may have been reported on traits that happened to be more heritable than traits for which the estimates of variance components were reported. Through theory, we show that such a discrepancy can arise when individual studies represent a mixture of traits that follow a pattern of $r_{MZ} > 2r_{DZ}$ and $r_{MZ} < 2r_{DZ}$ and where the choice of fitting a model that includes shared environment or non-additive genetic influences is based on the observed pattern of twin correlations (**Supplementary Note**). More specifically, because c^2 and non-additive genetic influences cannot be estimated simultaneously from twin correlations, an ACE model (for additive genetic (A), common environmental (C), and error or non-shared environmental (E) influences) is fitted to the data if $2r_{DZ} - r_{MZ} > 0$. In contrast, if $2r_{DZ} - r_{MZ} < 0$, an ADE model, including non-additive genetic (D) instead of common environmental (C) influences, is selected. This leads to sampling bias in estimating h^2 from the full model. We show (**Supplementary Table 26**) that such per-study choices cause bias and can lead to a 10% downward bias in the reported estimates of h^2 in comparison to those based on twin correlations, consistent with the observed discrepancy between our meta-analysis of variance component estimates calculated from twin correlations and the reported variance components.

Overall twin correlations imply a simple additive model

There may be many causes of similarities and differences within monozygotic and dizygotic twin pairs, and these are typically interpreted in terms of additive or non-additive genetic influences and shared or non-shared environmental influences[19]. Yet, there are essentially only two estimable and testable variance components of interest in the twin design. Therefore, inference from classical twin studies on all underlying, unobserved sources of variation that lead to the resemblance between relatives is limited. However, there are two simple and parsimonious hypotheses that can be tested across traits from estimated correlation coefficients for monozygotic twin pairs (r_{MZ}) and dizygotic twin pairs (r_{DZ}). The first is that the correlations for the monozygotic and dizygotic twin populations (ρ_{MZ} and ρ_{DZ}) are the same, consistent with twin resemblance being solely due to non-genetic factors. The second hypothesis involves a two-fold ratio of ρ_{MZ} to ρ_{DZ}, consistent with twin resemblance being solely due to additive genetic variation. Across-trait consistency with either of these hypotheses is not a proof of these simple models but would provide an extremely parsimonious model against which other experimental designs (for example, DNA sequence–based studies) should be tested. For the vast majority of traits (84%),

we found that monozygotic twin correlations were larger than dizygotic twin correlations. Using the weighted estimates of r_{MZ} and r_{DZ} across all traits, we showed that, on average, $2r_{DZ} - r_{MZ} = 0.042$ (s.e.m. = 0.007) (**Table 1**), which is very close to a twofold difference in the correlation of monozygotic twins relative to dizygotic twins (**Supplementary Figs. 11** and **12**). The proportion of single

Figure 3 Twin correlations for the top 20 most investigated specific traits by age and sex. Alc., alcohol; dis., disorders; depr., depressive; endocr., endocrine; imm., immunological; funct., functions; maint., maintenance; metab., metabolic; ment. beh., mental and behavioral; spec. personal., specific personality; temp. pers., temperament and personality; tob., tobacco; r, correlation; MZ, monozygotic twins; DZ, dizygotic twins; M, males; F, females; SS, same-sex pairs only; DOS, dizygotic opposite-sex pairs. Inclusion for the top 20 most investigated traits was conditional on the reporting of r_{MZ} and r_{DZ}. Empty cells denote insufficient information available to calculate weighted estimates; error bars, standard errors. We note that estimates and graphs for all specific traits are available from the online MaTCH webtool.

Table 2 Weighted means of twin correlations and proportion (π_0) of studies that are consistent with a model where trait resemblance is solely due to additive genetic variation for the main trait domains and the top 20 most investigated traits

	π_0		r_{MZ}			r_{DZ}		
	n traits	Estimate	Estimate (s.e.m.)	n traits	n pairs	Estimate (s.e.m.)	n traits	n pairs
All traits	5,185	0.69	0.636 (0.002)	9,568	2,563,628	0.339 (0.003)	5,220	2,606,252
General trait domains								
Activities	62	0.35	0.570 (0.019)	118	58,227	0.340 (0.022)	63	55,864
Cardiovascular	267	0.95	0.564 (0.008)	380	41,669	0.295 (0.010)	268	25,544
Cell	54	0.59	0.722 (0.022)	72	3,188	0.523 (0.043)	54	1,667
Cognitive	450	0.57	0.646 (0.007)	931	288,867	0.371 (0.010)	454	304,720
Dermatological	74	0.45	0.729 (0.025)	109	19,509	0.402 (0.017)	75	23,245
Ear, nose, throat	165	0.97	0.760 (0.013)	200	27,882	0.332 (0.015)	172	14,222
Endocrine	108	0.69	0.555 (0.017)	162	10,112	0.387 (0.022)	110	9,140
Environment	145	0.50	0.551 (0.014)	295	120,606	0.396 (0.017)	145	99,137
Gastrointestinal	32	0.59	0.551 (0.024)	64	10,982	0.274 (0.028)	39	28,431
Hematological	19	0.65	0.764 (0.023)	50	5,541	0.560 (0.032)	19	3,218
Immunological	230	0.67	0.608 (0.012)	280	18,051	0.357 (0.013)	231	36,075
Metabolic	464	0.60	0.746 (0.005)	912	210,189	0.405 (0.008)	464	197,921
Neurological	702	1.00	0.685 (0.005)	1,751	129,076	0.289 (0.006)	705	89,103
Nutritional	110	0.72	0.479 (0.016)	205	75,751	0.220 (0.015)	110	79,188
Ophthalmological	106	0.87	0.730 (0.017)	199	26,139	0.385 (0.017)	106	16,189
Psychiatric	1,778	0.62	0.552 (0.004)	2,865	1,232,382	0.306 (0.005)	1,781	1,374,817
Reproduction	16	0.44	0.767 (0.034)	34	12,130	0.333 (0.063)	16	27,879
Respiratory	125	0.74	0.697 (0.018)	184	34,443	0.325 (0.019)	127	51,150
Skeletal	190	0.51	0.830 (0.008)	395	111,282	0.504 (0.012)	191	113,080
Social interactions	24	0.63	0.338 (0.017)	146	43,501	0.267 (0.041)	24	22,764
Social values	45	0.69	0.489 (0.030)	120	52,492	0.414 (0.062)	45	28,071
Top 20 investigated traits for r_{MZ} and r_{DZ}								
Blood pressure functions	110	0.93	0.581 (0.010)	179	20,621	0.307 (0.013)	110	11,620
Conduct disorder	216	0.41	0.663 (0.009)	289	147,974	0.408 (0.010)	216	192,651
Depressive episode	115	0.60	0.454 (0.014)	173	98,315	0.253 (0.015)	115	121,936
Endocrine gland functions	92	0.72	0.538 (0.017)	139	8,533	0.382 (0.025)	92	7,295
Food	110	0.72	0.479 (0.016)	205	75,751	0.220 (0.015)	110	79,188
Functions of brain	594	0.99	0.676 (0.006)	1,010	69,722	0.287 (0.006)	594	58,951
General metabolic functions	219	0.69	0.682 (0.007)	462	62,108	0.371 (0.010)	219	58,338
Heart functions	140	1.00	0.529 (0.009)	174	15,070	0.268 (0.011)	140	11,109
Height	87	0.29	0.908 (0.005)	128	53,076	0.543 (0.008)	87	68,358
Higher-level cognitive functions	188	0.44	0.710 (0.009)	419	152,197	0.441 (0.016)	188	158,626
Hyperkinetic disorders	100	0.37	0.651 (0.013)	144	86,450	0.260 (0.016)	100	121,139
Immunological system functions	223	0.67	0.606 (0.012)	276	16,703	0.357 (0.013)	223	32,964
Mental and behavioral disorders due to the use of alcohol	100	0.36	0.630 (0.015)	158	94,477	0.409 (0.020)	101	94,196
Mental and behavioral disorders due to the use of tobacco	70	0.47	0.719 (0.016)	110	51,102	0.468 (0.022)	72	34,186
Other anxiety disorders	145	0.29	0.548 (0.013)	191	105,902	0.327 (0.016)	145	153,730
Specific personality disorders	140	0.93	0.448 (0.009)	162	41,460	0.225 (0.007)	140	33,681
Structure of the eyeball	86	0.91	0.735 (0.022)	121	19,276	0.370 (0.019)	86	13,580
Structure of mouth	117	0.89	0.819 (0.010)	127	7,769	0.399 (0.012)	119	8,493
Temperament and personality functions	568	0.84	0.470 (0.008)	1,134	334,190	0.234 (0.010)	568	296,114
Weight maintenance functions	215	0.48	0.810 (0.005)	391	141,152	0.437 (0.010)	215	134,867

General trait domain categories with <10 entries for π_0 were excluded. For definitions of abbreviations, see **Table 1**. Inclusion of the top 20 investigated traits was conditional on the reporting of r_{MZ} and r_{DZ}.

studies in which the pattern of twin correlations was consistent with the null hypothesis that $2r_{DZ} = r_{MZ}$ was 69%. This observed pattern of twin correlations is consistent with a simple and parsimonious underlying model of the absence of environmental effects shared by twin pairs and the presence of genetic effects that are entirely due to additive genetic variation (**Table 2**). This remarkable fitting of the data with a simple mode of family resemblance is inconsistent with the hypothesis that a substantial part of variation in human traits is due to shared environmental variation or to substantial non-additive genetic variation.

Most specific traits follow an additive genetic model

Although across all traits 69% of studies showed a pattern of monozygotic and dizygotic twin correlations consistent with an r_{MZ} that was exactly twice the r_{DZ}, this finding is not necessarily representative of the majority of studies in functional domains or for every specific trait (that is, at the ICD-10 or ICF subchapter level). We thus calculated the proportion of studies consistent with $2r_{DZ} = r_{MZ}$ within functional domains and for each specific trait and found that traits consistent with this hypothesis tended to cluster in specific functional domains (**Supplementary Tables 27–29**). A pattern of twin correlations

consistent with $2r_{DZ} = r_{MZ}$ was most prominent for traits included in the neurological, ear, nose and throat, cardiovascular and ophthalmological domains, with 99.5%, 97%, 95% and 87% of studies, respectively, being consistent with a model where all resemblance was entirely due to additive genetic variance. In only 3 of 28 general trait domains were most studies inconsistent with this model. These domains were activities (35%), reproduction (44%) and dermatological (45%) (**Table 2** and **Supplementary Table 27**). Of the 59 specific traits (ICD-10 or ICF subchapter classifications) for which we had sufficient information to calculate the proportion of studies consistent with $2r_{DZ} = r_{MZ}$, 21 traits showed a proportion less than 0.50, whereas for the remaining 38 traits the majority of individual studies were consistent with $2r_{DZ} = r_{MZ}$ (**Supplementary Table 29**). Of the top 20 most investigated specific traits, we found that for 12 traits the majority of individual studies were consistent with a model where variance was solely due to additive genetic variance and non-shared environmental variance, whereas the pattern of monozygotic and dizygotic twin correlations was inconsistent with this model for 8 traits, suggesting that, apart from additive genetic influences and non-shared environmental influences, either or both non-additive genetic influences and shared environmental influences are needed to explain the observed pattern of twin correlations (**Table 2**). These eight traits were conduct disorders, height, higher-level cognitive functions, hyperkinetic disorders, mental and behavioral disorders due to the use of alcohol, mental and behavioral disorders due to the use of tobacco, other anxiety disorders and weight maintenance functions. For all eight traits, meta-analyses on reported variance components resulted in a weighted estimate of reported shared environmental influences that was statistically different from zero (**Supplementary Table 21**). Comparison of weighted twin correlations for these specific traits resulted in positive estimates of $2r_{DZ} - r_{MZ}$, except for hyperkinetic disorders, where $2r_{DZ} - r_{MZ}$ was −0.130 (s.e.m. = 0.034) on the basis of 144 individual reports and 207,589 twin pairs, which suggests the influence of non-additive genetic variation for this trait or any other source of variation that leads to a disproportionate similarity among monozygotic twin pairs.

DISCUSSION

We have conducted a meta-analysis of virtually all twin studies published in the past 50 years, on a wide range of traits and reporting on more than 14 million twin pairs across 39 different countries. Our results provide compelling evidence that all human traits are heritable: not one trait had a weighted heritability estimate of zero. The relative influences of genes and environment are not randomly distributed across all traits but cluster in functional domains. In general, we showed that reported estimates of variance components from model-fitting can underestimate the true trait heritability, when compared with heritability based on twin correlations. Roughly two-thirds of traits show a pattern of monozygotic and dizygotic twin correlations that is consistent with a simple model whereby trait resemblance is solely due to additive genetic variation. This implies that, for the majority of complex traits, causal genetic variants can be detected using a simple additive genetic model.

Approximately one-third of traits did not follow the simple pattern of a twofold ratio of monozygotic to dizygotic correlations. For these traits, a simple additive genetic model does not sufficiently describe the population variance. An incorrect assumption about narrow-sense heritability (the proportion of total phenotypic variation due to additive genetic variation) can lead to a mismatch between the results from gene-finding studies and previous expectations. If the pattern of twin correlations is consistent with a substantial contribution from shared

environmental factors, as we find for conduct disorders, religion and spirituality, and education, then gene-mapping studies may yield disappointing results. If the cause of departure from a simple additive genetic model is the existence of non-additive genetic variation, as is, for example, suggested by the average twin correlations for recurrent depressive disorder, hyperkinetic disorders and atopic dermatitis, then it may be tempting to fit non-additive models in gene-mapping studies (for example, GWAS or sequencing studies). However, the statistical power of such scans is extremely low owing to the many non-additive models that can be fitted (for example, within-locus dominance versus between-locus additive-by-additive effects) and the penalty incurred by multiple testing. Our current results signal traits for which an additive model cannot be assumed. For most of these traits, dizygotic twin correlations are higher than half the monozygotic twin correlations, suggesting that shared environmental effects are causing the deviation from a simple additive genetic model. Yet, data from twin pairs only do not provide sufficient information to resolve the actual causes of deviation from a simple additive genetic model. More detailed studies may identify the likely causes of such deviation and may as such uncover epidemiological or biological factors that drive family resemblance. To make stronger inferences about the causes underlying resemblance between relatives for traits that deviate from the additive genetic model, additional data are required, for example, from large population samples with extensive phenotypic and DNA sequence information, detailed measures of environmental exposures and larger pedigrees including non-twin relationships.

We note that our inference is based on twin studies published between 1958 and 2012 and that it generally applies to complex traits but does not necessarily generalize to mendelian subtypes of traits. Most mendelian traits are rare in the population and are therefore not studied by researchers of twins because they cannot ascertain enough affected twin pairs to reliably estimate genetic parameters. In the rare case where sufficient numbers of affected twin pairs were available, the mendelian subtypes were analyzed together with the subtypes of the same trait that were due to common causes, as it was unknown whether the studied trait was a mendelian subtype. If the traits we studied were in fact a mix of mendelian and complex subtypes, our inference would be biased away from our main result because mendelian diseases tend to be dominant or recessive, not additive. In addition, there may be heterogeneity in measurement errors between studies for the same trait and between traits. A test-retest correlation would quantify measurement error when contrasted with a correlation between monozygotic twins, but few twin studies report such correlations in the same papers that estimate heritability.

Our results provide the most comprehensive empirical overview of the relative contributions of genes and environment to all human traits that have been studied in twins thus far, which can guide and serve as a reference for future gene-mapping efforts.

URLs. ICF classification, http://apps.who.int/classifications/icfbrowser/; ICD-10 classification, http://apps.who.int/classifications/icd10/browse/2010/en. The data used for this manuscript have been integrated in a web application, where user-specified selections of traits can be made to apply the analyses presented in this work. The web application is called MaTCH (Meta-analysis of Twin Correlations and Heritability) and is accessible via http://match.ctglab.nl/; Gephi, http://gephi.github.io/.

METHODS

Methods and any associated references are available in the online version of the paper.

AUTHOR CONTRIBUTIONS

D.P., B.B., P.F.S. and P.M.V. performed the analyses. D.P. conceived the study. D.P., T.J.C.P. and P.M.V. designed the study. T.J.C.P. and D.P. collected and entered the data. D.P. and P.F.S. categorized traits according to standard classifications. A.v.B. and C.A.d.L. checked data entries, and checked and wrote statistical scripts. A.v.B. designed and programmed the webtool. D.P., T.J.C.P. and P.M.V. wrote the manuscript. All authors discussed the results and commented on the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Moore, J.H. Analysis of gene-gene interactions. *Curr. Protoc. Hum. Genet.* **Chapter 1**, Unit 1.14 (2004).
2. Hill, W.G., Goddard, M.E. & Visscher, P.M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* **4**, e1000008 (2008).
3. Traynor, B.J. & Singleton, A.B. Nature versus nurture: death of a dogma, and the road ahead. *Neuron* **68**, 196–200 (2010).
4. Zuk, O., Hechter, E., Sunyaev, S.R. & Lander, E.S. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. USA* **109**, 1193–1198 (2012).
5. Phillips, P.C. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* **9**, 855–867 (2008).
6. Visscher, P.M., Brown, M.A., McCarthy, M.I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
7. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
8. Stranger, B.E., Stahl, E.A. & Raj, T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* **187**, 367–383 (2011).
9. Maher, B. Personal genomes: the case of the missing heritability. *Nature* **456**, 18–21 (2008).
10. Eichler, E.E. *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* **11**, 446–450 (2010).
11. Nelson, R.M., Pettersson, M.E. & Carlborg, Ö. A century after Fisher: time for a new paradigm in quantitative genetics. *Trends Genet.* **29**, 669–676 (2013).
12. Barker, J.S. Inter-locus interactions: a review of experimental evidence. *Theor. Popul. Biol.* **16**, 323–346 (1979).
13. Cockerham, C.C. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**, 859–882 (1954).
14. Cockerham, C.C. in *Statistical Genetics and Plant Breeding* 53–94 (Nat. Acad. Sci. Nat. Res. Council Publ., 1963).
15. Kempthorne, O. On the covariances between relatives under selfing with general epistacy. *Proc. R. Soc. Lond. B Biol. Sci.* **145**, 100–108 (1956).
16. Crow, J.F. & Kimura, M. *An Introduction To Population Genetics Theory* (Harper and Row, 1970).
17. Carlborg, O. & Haley, C.S. Epistasis: too often neglected in complex trait studies? *Nat. Rev. Genet.* **5**, 618–625 (2004).
18. Falconer, D.S. & Mackay, T.F.C. *Quantitative Genetics* (Longman Group, 1996).
19. Lynch, M. & Walsch, B. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, 1998).

ONLINE METHODS

Identifying relevant studies. We searched PubMed for studies published between 1 January 1900 and 31 December 2012 that provided twin correlations, concordance rates, and a heritability estimate (h^2) or an estimate of shared environmental influences (c^2), using monozygotic and dizygotic twins. The following search term was used

("English"[Language] AND ("1900/01/01"[Date - Publication]: "2012/12/31" [Date - Publication]) AND twin AND "journal article"[Publication Type] AND "humans"[Filter] AND (heritability[Title/Abstract] OR "genetic influence"[Title/Abstract] OR "environmental influence"[Title/Abstract] OR "genetic factors"[Title/Abstract] OR "environmental factors"[Title/Abstract]) AND "journal article"[Publication Type]) NOT review[Title] NOT review [Publication Type])

The search was run on 31 January 2013 and again on 29 April 2013, which yielded an additional 44 publications, with the difference likely due to keywords or tags that had been added to publications in the intermittent period.

The last PubMed search yielded 4,388 unique studies. From these, we deleted studies that were not relevant for the current purpose using the following exclusion criteria: (i) studies with only monozygotic twins available; (ii) studies with no heritability estimate, twin correlations or concordances available; (iii) review studies; (iv) meta-analyses; and (v) multivariate studies that provided information on completely overlapping traits and samples with previously published univariate studies. Some studies investigating h^2 for the brain (for example, voxel-based brain measures) were not included for practical purposes. These studies typically presented their results in graphs with color-coded point estimates of heritability mapped onto the brain. Such estimates could not be quantified, and these studies were thus not included.

From the remaining 2,748 studies, we were able to retrieve the full text from all but 5 papers (99.8%). Of the studies without full-text availability, we included relevant information based on the abstract. An overview of authors and journals and a full reference list of all 2,748 studies are provided in **Supplementary Tables 30–32**.

Primary information obtained from each study. From every study, we retrieved basic information on the PubMed ID, the authors, the trait as named in the study and the year of publication. In addition, the following information was retrieved:

- Country of origin of the study population. We used standard ISO country names, and where possible data entry was done separately for each country investigated in the study.
- Age group of the study population. The study population was classified into four age cohorts on the basis of the average age of the included sample: age >0 and <12; age ≥12 and <18; age ≥18 and <65; and age ≥65.
- Monozygotic and dizygotic twin correlations. Twin correlations were entered as provided in the study and could be calculated as intraclass, Pearson, polychoric or tetrachoric correlations or on the basis of least-squares or maximum-likelihood estimates. When available, we entered the twin correlations separately for males and females (i.e., monozygotic male (MZM), monozygotic female (MZF), dizygotic male (DZM), dizygotic female (DZF) and dizygotic opposite-sex (DOS) pairs). If correlations were not available for males and females separately, we entered the MZ and DZ correlations, i.e., the correlations based on both sexes. In cases where it was clear that the dizygotic correlation was based on same-sex twins only, we entered the dizygotic same-sex (DZSS) correlation.
- Estimates of heritability (h^2) and shared environmental component (c^2), under the full ACE (or ADE) model. We entered 'h^2_FULL' and 'c^2_FULL', on the basis of estimates under the full ACE (including additive genetic and shared and non-shared environmental influences) or ADE (including additive and non-additive genetic and non-shared environmental influences) model. When an ACE model was fitted, the estimate for A was entered in 'h^2_FULL' and the estimate for C was entered in 'c^2_FULL'. When an ADE model was fitted, the estimates of A and D were summed and entered for 'h^2_FULL' and zero was entered for 'c^2_FULL'. When estimates were provided separately for males and females, they were entered

separately. In the case of multivariate analyses, univariate estimates were always preferred to allow comparison across studies.

- Estimates of heritability (h^2) and shared environmental component (c^2), under the best-fitting ACE (or ADE) model. We entered 'h^2_BEST' and 'c^2_BEST', on the basis of estimates under the best-fitting ACE or ADE model as provided in the study. When an ACE model was the best-fitting model, the estimate for A was entered in 'h^2_BEST' and the estimate for C was entered in 'c^2_BEST'. When an ADE model was the best-fitting model, the estimates of A and D were summed and entered for 'h^2_BEST' and zero was entered for 'c^2_BEST'. When estimates were provided separately for males and females, they were entered separately. In the case of multivariate analyses, univariate estimates were always preferred to allow comparison across studies. In cases where estimates for the best-fitting model were not directly provided but information available in the paper indicated that the best-fitting model was AE (or CE or E), we entered zero for 'c^2_BEST' and missing for 'h^2_BEST' (when the best-fitting model was an AE model), missing for 'c^2_BEST' and zero for 'h^2_BEST' (when the best-fitting model was a CE model), and zero for 'c^2_BEST' and zero for 'h^2_BEST' (when the best-fitting model was described to be an E model).
- The total number of twin pairs as used for each entered estimate.
- Whether the study was a classical twin study. All 2,748 studies in the database included monozygotic and dizygotic twins. However, a classical twin study was defined as a study that involved only reared-together monozygotic and dizygotic twins. From studies that included siblings, extended families, adoptees or reared-apart twins, only estimates based on the reared-together twin sample were used for the meta-analyses. Most of the non-classical twin studies did provide twin correlations for the classical twin design and were thus included in the meta-analysis for twin correlations. When A and C estimates were based on extended twin designs, they were excluded from the meta-analyses.
- The method used for estimating the variance components. We entered the statistical method used for estimating the variance components, which included, for example, ANOVA, Bayesian, maximum-likelihood (ML), DeFries-Fulker regression, least-squares (LS) or intrapair differences. We also listed a dichotomized version of this indicating whether the method used was 'ML or LS' or 'not ML or LS', for all other methods. In the meta-analyses for h^2 and c^2 estimates, only those based on maximum likelihood or least squares were included.
- Whether the trait was dichotomous or continuous. Traits measured as 0 or 1, as well as traits measured on a quantitative scale but dichotomized before analysis, were listed as dichotomous. All other traits, including ordinal traits, were listed as 'not dichotomous' and treated as continuous.
- Whether the study involved ascertainment for the trait. When the trait under investigation was the same trait that was used to select probands, the study was listed as 'ascertained'.
- Number of concordant and discordant pairs. In cases of dichotomous traits, the total numbers of pairs for discordant and concordant affected pairs were entered separately for each zygosity. In cases of dichotomous traits that were not ascertained, the number of concordant unaffected pairs was also entered.
- Prevalence. In cases of dichotomous traits, the population prevalence, separately for monozygotic and dizygotic twins when available, was entered. Prevalence was based on what was provided in the study or was calculated using $(2c + d)/2n$, where c is the number of concordant affected pairs, d is the number of discordant pairs and n is the total number of pairs in non-ascertained traits.

Thus, provided that there was availability, the statistics in **Supplementary Table 1** were obtained for each trait reported on in every study. When the five basic twin correlations were available (r_{MZM}, r_{MZF}, r_{DZM}, r_{DZF} and r_{DOS}), we calculated r_{MZ}, r_{DZSS} and r_{DZ}, using the weighted average via Fisher z transformation and using sample size as weights. In situations where r_{MZM} (or r_{MZF}) was exactly 1 (or −1), we subtracted (or added in the case of −1) 0.00001 to the correlation to ensure non-problematic Fisher z transformation. Sample sizes of MZ, DZSS and DZ were obtained by summing the sample

sizes of MZM and MZF, of DZM and DZF, and of DZM, DZF and DOS, respectively. Estimates of h^2 and c^2 were calculated across sex as the n-weighted average across the separate male and female estimates, when available. For the number of concordant and discordant pairs, MZ, DZ and DZSS were calculated on the basis of the numbers available for MZM, MZF, DZM, DZF and DOS. Prevalences for pooled entries were calculated as an n-weighted average.

Data entry checks. Studies were entered and cross-checked for obvious typos by T.J.C.P. and D.P. After initial data entry and initial cross-checking, all data points were manually checked (D.P.) by looking up the entered values in the original paper. In addition, automatic checks were run (D.P. and B.B.) to identify typos, strange outliers or obvious mistakes. These checks included: (i) identifying highly unlikely values (clear typos, for example, correlation of 120); (ii) testing whether the sum of h^2 and c^2 was <100; (iii) testing for strange discrepancies between estimates from the full and best-fitting models; and (iv) checking for outliers on the basis of extreme sample size and extreme χ^2 values for rejecting the null hypothesis that either $2 \times (r_{MZ} - r_{DZ})$ or $2 \times r_{DZ} - r_{MZ}$ is equal to zero.

Classification of traits. After data entry, all traits were manually classified using the ICF. The ICF is the framework of the World Health Organization (WHO) for health and disability and provides the conceptual basis for the definition, measurement and policy formulations for health and disability. It is a universal classification of disability and health for use in health and health-related sectors. ICF belongs to the WHO family of international classifications, the best-known member of which is the ICD-10. ICD-10 provides an etiological framework for the classification of diseases, disorders and other health conditions, whereas ICF classifies functioning and disability associated with health conditions. The ICD-10 and ICF are therefore complementary (see URLs).

Most traits investigated in twin studies concern healthy functioning (for example, cognitive function, social attitudes, body height and personality) and were classified according to ICF. In cases where the studied traits were diseases or symptoms of disease, ICD-10 was used. Traits were given two hierarchical classifications corresponding to the ICF or ICD-10 hierarchical structure, using the chapter structure (for example, b1) and the level directly under the chapter (for example, b110), which corresponds to the code for the actual disease (ICD-10) or trait (ICF).

Six new classes at the chapter level and 17 new classes at the subchapter level were created to accommodate traits that could not be classified under either ICF or ICD-10. For the chapter level, the created classes were cell, function of DNA, functions of the nervous system, medication effects, mortality and structure of DNA. For the subchapter level, the classes created were all-cause mortality, cell cycle, cell growth, diazepam effects, expression, function of brain, gene expression, height, methylation, mortality from heart disease, mtDNA, physical appearance, receptor binding, sister chromatid exchange, structure of DNA, telomeres and X inactivation.

In addition to the two standard ICF or ICD-10 classification levels, we added a general classification of functional trait domains. We thus classified all traits using a 3-level scheme that included 28 broad, functional domains, 54 ICF or ICD-10 chapter-level classes and 313 subchapter-level classes. A small proportion of studied traits (<0.1%) could not be classified meaningfully on the chapter level (two traits) or the subchapter level (three traits). There were 326 unique combinations across the 3 levels of trait categorization (**Supplementary Table 33**). All analyses were conducted on all entries of each of the three levels of classification. In addition, we analyzed all traits together. Although this is unspecific in terms of diseases or traits, it provides a general overview of the relationship between monozygotic and dizygotic twin correlations and shows general patterns of, for example, sex and age differences. The most specific level was the subchapter level, which was the actual ICD-10 diagnosis or a similar ICF classification for normal functioning, reflecting specific traits such as cleft lip, hyperkinetic disorders or higher-level cognitive function. As researchers do not necessarily adhere to the ICD-10 or ICF trait nomenclature, traits with the same subchapter classification could have different trait names in the original study: for example, for higher-level cognitive function, the original studies included the trait names total IQ score, cognitive ability, intelligence or 'g'.

Tests for publication bias. Publication bias can occur when studies that report relatively large heritability estimates or high twin correlations are more likely to be submitted and/or accepted for publication than studies that report more modest effects. Such a publication bias would lead to an overestimation of the true twin correlations or the true heritability and environmental estimates. We used several standard statistical tools to aid in identifying and quantifying possible publication bias, including inspection of funnel plots, Begg and Mazumdar's test[20], Egger's regression test[21] and Rosenthal's fail-safe N[22].

Meta-analysis methods of twin correlations and variance components. We used the DerSimonian-Laird (DSL) random-effects meta-analytical approach with correlation coefficients as effect sizes, as described by Schulze[23] and implemented in the R package metacor. This function transforms a correlation to its Fisher z value with corresponding standard error before the meta-analysis. This method is preferred over conducting a meta-analysis directly on the correlations because the standard error of a twin correlation is a function of not only sample size but also the correlation itself, with larger correlations having a smaller standard error. This can cause problems in a meta-analysis, as it would lead to the larger correlations appearing more precise and being assigned more weight in the analysis, irrespective of sample size. To avoid this problem, the DSL method transforms correlations to the Fisher's z metric, whose standard error is determined solely by sample size. All n-weighted computations were thus performed using Fisher's z metric, and the results were converted back to correlations for interpretation.

The random-effects approach allows for heterogeneity of the true twin correlations across different studies. That is, rather than assuming that there is one true level of twin correlation, the random-effects model allows a distribution of true correlations. The combined effect of the random-effects model represents the mean of the population of true correlations. For computational reasons, correlations of −1 and 1 were converted to −0.99999 and 0.99999 before meta-analysis. We set a threshold of at least five pairs of twins available per estimate and at least two studies available per category. Meta-analyses were conducted for each category of all three levels of classification.

We note that twin samples used in different publications were not independent. For example, studies using Australian twins are predominantly based on twins from the Australian Twin Registry. These studies sometimes include different subsamples but may also include completely overlapping samples used to investigate different traits. As participants are anonymous, it is not possible to determine the extent of overlap in the studies included in our analyses. We thus assumed independency of samples in the meta-analyses. This assumption leads to an underestimation of the variance of weighted estimates and an overestimation of their precision. We expect that the dependency of study samples is lowest at the specific level of the ICD-10 or ICF subchapters and highest for the general functional domains.

Meta-analysis for dichotomous, non-ascertained traits. In the DSL random-effects model, the standard error of a correlation is calculated on the basis of the provided n (pairs). The estimated standard errors for continuous traits are correct, but for dichotomous traits the resulting standard error is incorrect. That is because twin correlations for non-ascertained, dichotomous traits are typically based on three categories of pairs: concordant, unaffected (CON⁻), discordant (DIS) and concordant, affected (CON⁺) pairs. Whereas the total number of participating pairs is the sum of these, the information that determines the twin correlation and its significance is mostly derived from the latter two categories. For non-ascertained, dichotomous traits, we calculated the study-specific tetrachoric twin correlation on the basis of the contingency table (i.e., CON⁻, DIS and CON⁺ pairs), under the assumption that the dichotomous traits represent latent variables that follow a bivariate normal distribution[24]. We used a maximum-likelihood estimator described by Olsson[25], implemented in the R function polycor, to calculate the study-specific twin correlation and its standard error. As our meta-analysis required twin correlation and sample size (not standard error) as input and we wanted to be able to pool across continuous and dichotomous traits, we calculated the 'effective' number of pairs on the basis of the obtained standard error.

The effective number of pairs was defined as the number of pairs that produced the exact same standard error within the DSL meta-analyses as the standard error obtained from the contingency table.

Meta-analysis for dichotomous, ascertained traits. For ascertained traits, it was not possible to calculate the twin correlations and standard errors on the basis of the contingency table, as the traits included only pairs with at least one proband. Without information on the number of concordant, unaffected pairs, the prevalence of the affected status would be required to calculate a twin correlation. We used the algorithms derived from Falconer[26] and Smith[27]. Again, for practical purposes, calculated standard errors were transformed to an effective number of pairs for use in the DSL meta-analysis.

Proportion of studies consistent with specific hypotheses. We estimated the proportion of studies that were consistent with H_0: $2 \times (r_{MZ} - r_{DZ}) = 0$ ($\pi_0(h)$) and the proportion of observations consistent with H_0: $2 \times r_{DZ} - r_{MZ} = 0$ ($\pi_0(c)$), using the Jiang and Doerge method[28], as well as the q-value method[29].

Authorship network analysis. We used the approach more fully described previously[30]. Briefly, we retrieved from PubMed the full Medline listing for all twin studies included in this meta-analysis using NCBI eutils. The output was parsed to capture the names of all authors. The twin study author list was manually reviewed to resolve clear inconsistencies in the spelling of the names of authors between publications. Gephi was used to construct a network to understand twin study publication patterns. For clarity, we removed individuals who had published only one paper (i.e., we required authorship on ≥2 papers). The substructure of the network was investigated by estimating community membership modules using the Louvain method[31] implemented in Gephi.

20. Begg, C.B. & Mazumdar, M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* **50**, 1088–1101 (1994).
21. Egger, M., Davey Smith, G., Schneider, M. & Minder, C. Bias in meta-analysis detected by a simple, graphical test. *Br. Med. J.* **315**, 629–634 (1997).
22. Rosenthal, R. The file drawer problem and tolerance for null results. *Psychol. Bull.* **86**, 91–106 (1979).
23. Schulze, R. *Meta-Analysis: A Comparison Of Approaches* (Hogrefe & Huber, 2004).
24. Drasgow, F. in *Encyclopedia of Statistical Sciences* (eds. Kotz, S., Read, C.B., Balakrishnan, N. & Vidakovic, B.) Vol. 7, 68–74 (John Wiley & Sons, 2006).
25. Olsson, U. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika* **44**, 443–460 (1979).
26. Falconer, D.S. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann. Hum. Genet.* **29**, 51–76 (1965).
27. Smith, C. Concordance in twins: methods and interpretation. *Am. J. Hum. Genet.* **26**, 454–466 (1974).
28. Jiang, H. & Doerge, R.W. Estimating the proportion of true null hypotheses for multiple comparisons. *Cancer Inform.* **6**, 25–32 (2008).
29. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440–9445 (2003).
30. Bulik-Sullivan, B.K. & Sullivan, P.F. The authorship network of genome-wide association studies. *Nat. Genet.* **44**, 113 (2012).
31. Blondel, V., Guillaume, J., Lambiotte, R. & Lefebvre, E. Fast unfolding of community hierarchies in large networks. *J. Stat. Mech.* **10**, P10008 (2008).